Representation Learning for Dynamic Graphs

M. Sc. Seminar

Erfan Loghmani
Supervisor: Dr. MohammadAmin Fazli

Sharif University of Technology

Motivation

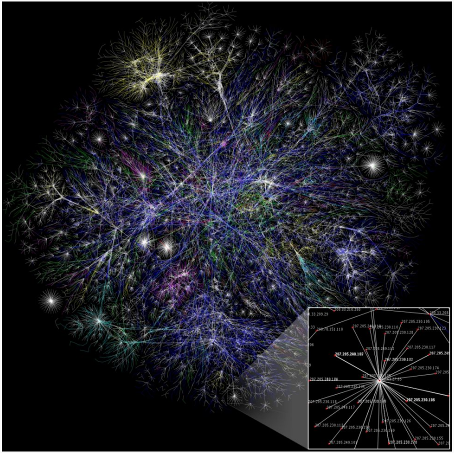


Figure 1: Partial map of the Internet based on the Jan. 15, 2005, data found on opte.org, The Opte Project

Motivation (contd.)

Problems on graphs:

- Node classification
- Link prediction
- Community detection
- Graph similarity

Motivation (contd.)

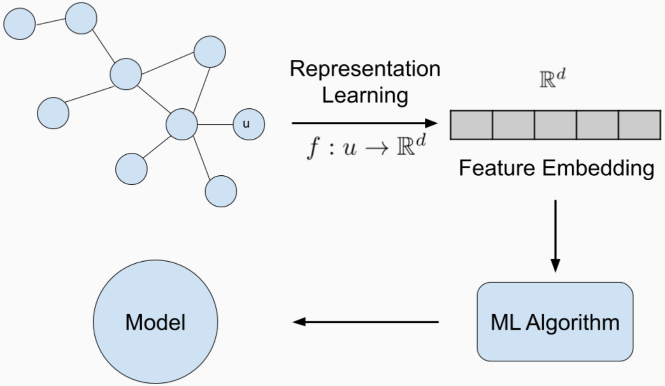


Figure 2: Representation learning position in machine learning pipeline.

Motivation (contd.)

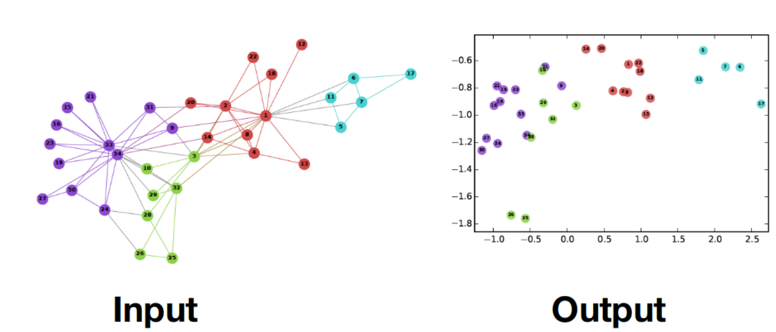


Figure 3: Example of representation learning for Zachary's karate club network. Perozzi et. al. Proceedings of the 20th ACM SIGKDD 2014

Table of contents

Introduction

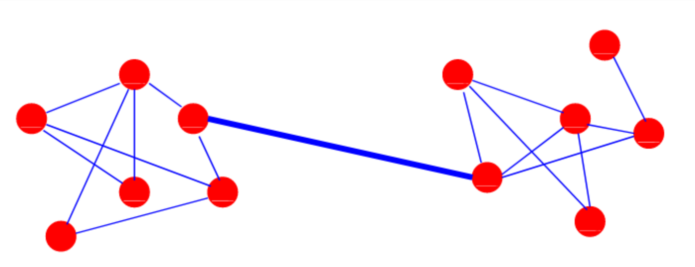


Figure 4: Girvan Newman one of the classic methods of community detection uses betweenness centrality to remove edges between communities. Fortunato et. al. 2009, in Encyclopedia of Complexity and Systems Science

Classic methods drawbacks

- Does not model problem complexity well
- Most of them can't use other node features
- Running time can grow fast

Dynamic graphs

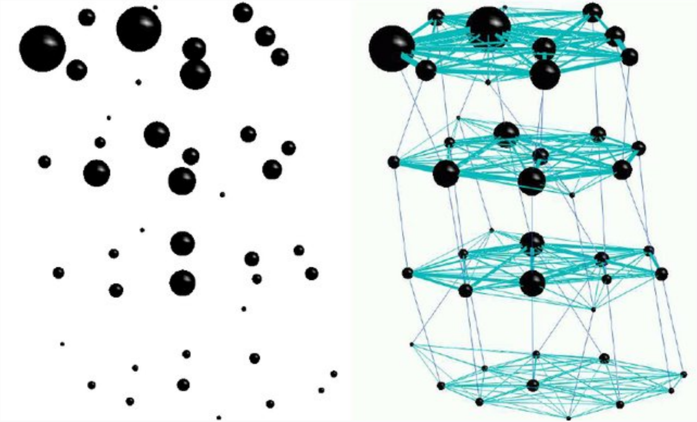


Figure 5: Graphs changing over time. Kobourov, S. G. (2012). Spring embedders and force directed graph drawing algorithms. arXiv preprint

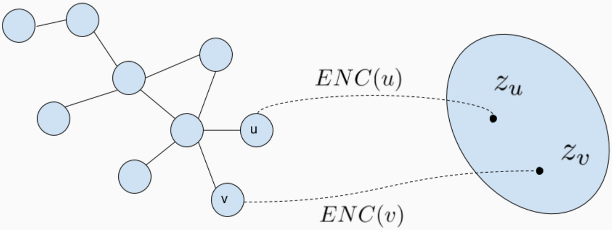
Using static methods for dynamic graphs

Using static methods on dynamic graph:

- increases running time as we should repeat the whole algorithm.
- does not model time dependent behaviour of data.
- may result unwanted big change in embeddings of same node in time.

Representation learning methods for static graphs

Encoder-Decoder approach [5]



$$Goal: \ similarity(u,v) \approx DEC(z_u, z_v) = z_u^T z_v$$

Figure 6: Encoder-Decoder approach schema. Hamilton et. al. "Representation learning on networks", WWW-18 Tutorial, 2018

Shallow approaches [5]

These methods simply define *ENC* function as

$$ENC(v_i) = \mathbf{Z}v_i \tag{1}$$

Where $\mathbf{Z} \in \mathbb{R}^{d \times |\mathcal{V}|}$ and v_i is one-hot encoding for node i.

Based on how we define similarity measure on graph we will get

- Adjacency based methods e.g. Graph Factorization [1], GraRep [2]
- Random walk methods e.g. Deep Walk [9], node2vec [3]

Graph neural networks [5]

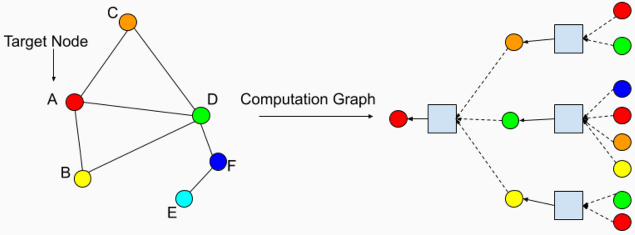


Figure 7: Graph neural networks schema. Hamilton et. al. ”Representation learning on networks”, WWW-18 Tutorial, 2018

Graph neural networks variants

- Graph Convolutional Networks Kipf et. al. 2016 [6]
- Graph Attention Networks Veličković et. al. 2017 [12]
- Gated Graph Neural Networks Li et. al. 2015 [8]
- GraphSAGE Hamilton et. al. 2017 [4]

Representation learning methods for dynamic graphs

Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks [7]

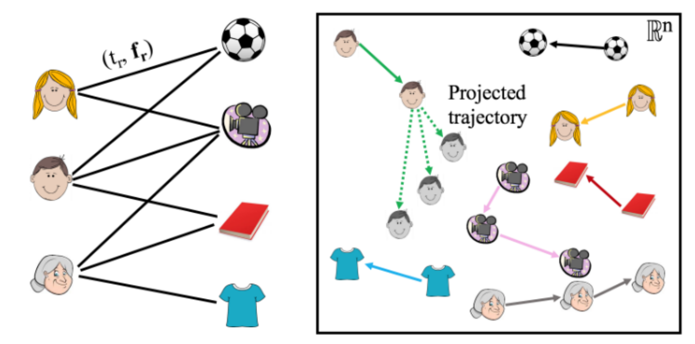


Figure 8: Problem settings and intuition. Kumar et. al. "Predicting dynamic embedding trajectory in temporal interaction networks." Proceedings of the 25th ACM SIGKDD 2019

Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks (contd.)

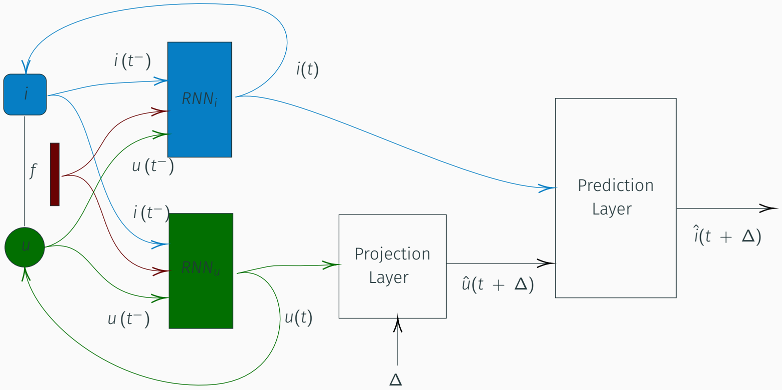


Figure 9: Model overview

Research direction

Cold start problem

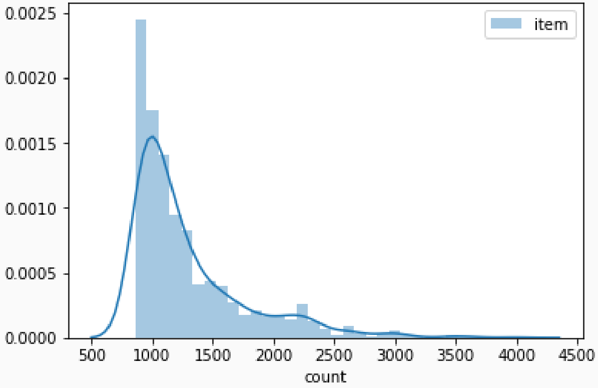


Figure 10: Histogram of number of interactions with each item in LastFM dataset used by Kumar et. al.

Simple projection layer

In JODIE projection layer is modeled as

$$\hat{u}(t+\Delta) = (1 + W_p\Delta) * u(t) \tag{2}$$

which $*$ is element-wise product.

Simple projection layer (contd.)

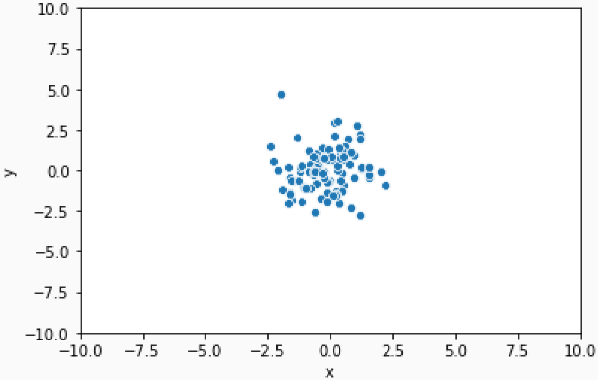


Figure 11: Representation projection in a simple 2d exampe for $\Delta = 0$

Simple projection layer (contd.)

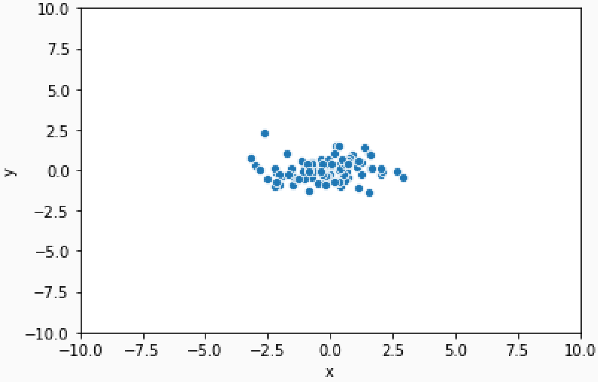


Figure 12: Representation projection in a simple 2d exampe for $\Delta = 0.2$

Simple projection layer (contd.)

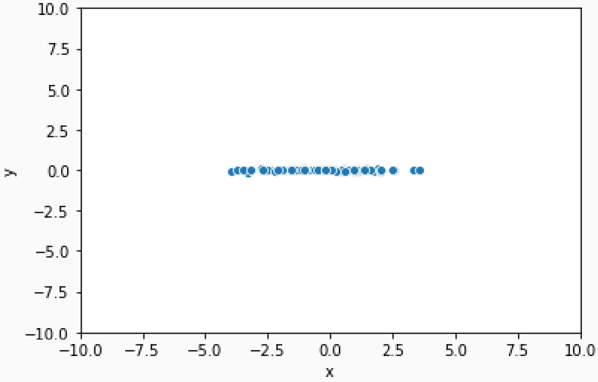


Figure 13: Representation projection in a simple 2d exampe for $\Delta = 0.4$

Other recommendations

- Prediction model now only predicts one point in embedding space. What about a probability distribution?
- Extend the model for one component graphs like social networks
- Use notations from social network theory like strategic network formation to better model individual behaviour and incentives.

Schedule

Table 1: Research schedule

Investigate on cold start problem	18th Bahman 1398
Try other methods for projection and prediction	23th Bahman 1398
Investigate on strategic network formation methods and their application in this problem	30th Farvardin 1399
Conclude and choose a direction	30th Ordibehesht 1399
Run experiments and evaluate different methods	17th Mordad 1399
Finalize thesis and present it	21th Shahrivar 1399

Questions?

A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, and A. J. Smola.
Distributed large-scale natural graph factorization.
In *Proceedings of the 22nd international conference on World Wide Web*, pages 37–48. ACM, 2013.

S. Cao, W. Lu, and Q. Xu.
Grarep: Learning graph representations with global structural information.
In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 891–900. ACM, 2015.

A. Grover and J. Leskovec.
node2vec: Scalable feature learning for networks.
In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.

W. Hamilton, Z. Ying, and J. Leskovec.
Inductive representation learning on large graphs.
In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

W. L. Hamilton, R. Ying, and J. Leskovec.
Representation learning on graphs: Methods and applications.
arXiv preprint arXiv:1709.05584, 2017.

T. N. Kipf and M. Welling.
Semi-supervised classification with graph convolutional networks.
arXiv preprint arXiv:1609.02907, 2016.

S. Kumar, X. Zhang, and J. Leskovec.
Predicting dynamic embedding trajectory in temporal interaction networks.
In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1269–1278, 2019.

Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel.
Gated graph sequence neural networks.
arXiv preprint arXiv:1511.05493, 2015.

B. Perozzi, R. Al-Rfou, and S. Skiena.
Deepwalk: Online learning of social representations.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

A. Sankar, Y. Wu, L. Gou, W. Zhang, and H. Yang.
Dynamic graph representation learning via self-attention networks.
arXiv preprint arXiv:1812.09430, 2018.

R. Trivedi, M. Farajtabar, P. Biswal, and H. Zha.
Dyrep: Learning representations over dynamic graphs.
In *International Conference on Learning Representations*, 2019.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio.
Graph attention networks.
arXiv preprint arXiv:1710.10903, 2017.

Dyrep: Representation Learning over Dynamic Graphs [11]

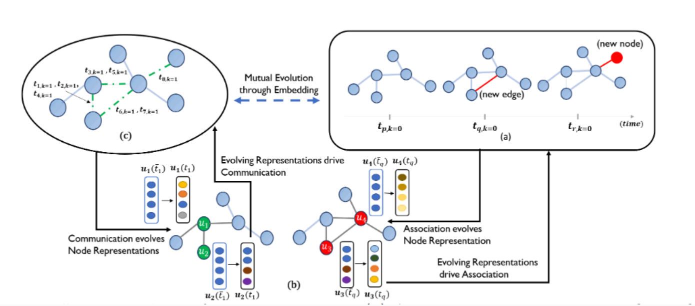


Figure 14: Dyrep general view. Trivedi, et al. "DyRep: Learning Representations over Dynamic Graphs." International Conference on Learning Representations. 2019

Dynamic Graph Representation Learning via Self-Attention Networks [10]

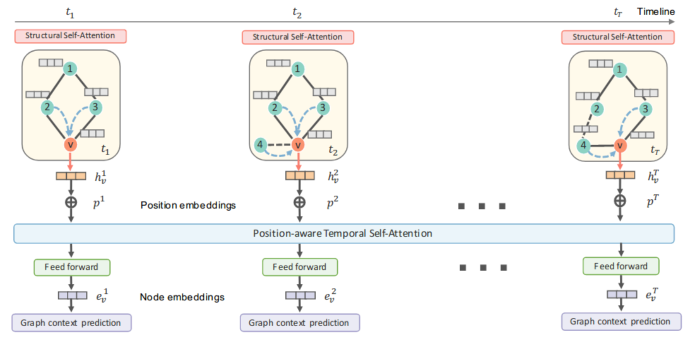


Figure 15: Model overview. Sankar et al. "Dynamic graph representation learning via self-attention networks." arXiv preprint 2018

JODIE formulations

Update:

$$
\begin{aligned}
u(t) &= \sigma(W_1^u u(t^-) + W_2^u i(t^-) + W_3^u f + W_4^u \Delta_u) \\
i(t) &= \sigma(W_1^i i(t^-) + W_2^i u(t^-) + W_3^i f + W_4^i \Delta_i)
\end{aligned} \tag{3}
$$

Projection:

$$\hat{u}(t+\Delta) = (1 + W_p \Delta) * u(t) \tag{4}$$

Prediction:

$$\hat{\bar{i}}(t+\Delta) = W_1 \hat{u}(t+\Delta) + W_2 \bar{u} + W_3 i(t+\Delta^-) + W_4 \bar{i} + B \tag{5}$$

loss:

$$
\begin{aligned}
loss = \sum_{(u,j,t,f)\in\mathcal{S}} & ||\hat{\bar{j}}(t) - [\bar{j}, j(t^-)]||_2 \\
& + \lambda_U ||u(t) - u(t^-)||_2 + \lambda_I ||j(t) - j(t^-)||_2
\end{aligned} \tag{6}
$$